# Constitutional Factors in the Production of Artificial Protein Fiber\*

Sam R. Hoover, Elsie L. Kokes, and Robert F. Peterson

Eastern Regional Research Laboratory.† Philadelphia, Pennsylvania

#### Abstract

Recent developments in the chemistry of artificial protein fiber are reviewed, with emphasis on the differences in amino acid content of the various proteins. New experimental results of thermoelastic analysis of protein fibers and determinations of peptide chain length in proteins by end-group analysis are presented.

A DISCUSSION of recent advances in artificial protein fibers—azlons—must necessarily cover almost all the work done in this field, for in the long history of textile fibers these have received serious attention for only about a decade. In this period many millions of pounds of such fibers have been

produced, and, more significantly, active research is being carried out today in a large number of laboratories. The early patents and publications on artificial protein fibers were reviewed in 1945 by Peterson, Caldwell. Hipp, Hellbach, and Jackson [63]. In recent years certain investigations have thrown light on the basic chemistry of protein fibers. The Regional Research Laboratories of the Department of Agriculture have emphasized these basic aspects of the problem. This discussion therefore will stress both the published and as yet unpublished work of these Laboratories.

<sup>\*</sup> Presented at the 18th Annual Meeting of the Textile Research Institute, New York, November, 1947, under the title "Research in Artificial Protein Fibers."

<sup>\*†</sup>One of the laboratories of the Bureau of Agricultural Industrial Chemistry, Agricultural Research Administration, United States Department of Agriculture.

The work we intend to describe falls rather naturally into four divisions: (a) the internal organization or physical structure of the fiber, (b) the chemical composition, (c) the reactivity of the constituent groups, and (d) studies of chain length in proteins by end-group determinations.

#### Internal Organization

The concept of an oriented, crystalline structure in fibrous materials was developed primarily in the study of cellulose and synthetic linear polymers. Meyer and Mark [57]. Astbury [2], and many others showed that this concept could be applied to the natural protein fibers—silk, wool, and collagen. Therefore, the conversion of a globular protein, edestin from hempseed, to the linear  $\beta$ -keratin structure characteristic of stretched wool by Astbury, Dickinson, and Bailey [4] was a discovery of great interest. They east films from solutions of edestin in urea and calcium chloride, whose denaturing action on the soluble globular proteins was already known, and, after washing out the reagent, obtained a relatively strong, elastic film which had an x-ray pattern similar to that of unoriented keratin. When this material was stretched 200 to 300 percent an oriented  $\beta$ -keratin pattern was obtained. They considered this to be a demonstration that denaturation involves the liberation of peptide chains which on coagulation aggregate into parallel bundles like those in  $\beta$ -keratin. Later, Lungren and O'Connell [48] produced relatively strong fibers from ovalbumin by spinning a solution of this protein in Nacconol,\* an alkyl-aryl sulfonate, into a precipitating bath and subsequently removing the denaturing agent by acctone-water extraction. Senti, Copley, and Nutting [71] showed that these were special cases of a general phenomenon. A number of proteins were crystallized, in the high polymer sense, by moist heat, phenols, acids, alcohols, amides, and other denaturing agents. These crystallites were then oriented by stretching, and a  $\beta$ -keratin pattern with essentially identical spacings resulted in each case. The resultant fibers had appreciably greater tenacity than the unoriented material, and possessed a truly fibrous structure.

Nutting, Halwer, Copley, and Senti [59] made a further valuable contribution by a study of the overall tensile properties—tenacity, stiffness, flexibility, toughness, and elasticity—of both natural and arti-

ficial protein fibers as they are related to the molecular configuration. A striking correlation of properties and configuration was found for keratins, silk, collagen, and ovalbumin. Ovalbumin was used as a model of the readily denaturable proteins. The high tenacity of the oriented fibrous form was demonstrated, but unfavorable changes in toughness, flexibility, and stiffness more than offset the advantage gained. Silk fibroin was exceptional in that it had good flexibility in its natural highly ordered state. The low flexibility of oriented ovalbumin was considered to indicate a low effective chain length by analogy with the well-established effect of chain length on flexibility in other high polymers. therefore reach the provisional conclusion that a highly oriented  $\beta$ -keratin structure possesses a high tenacity but is deficient in other essential properties.

A related but entirely independent approach to the study of the internal organization of fibrous materials has been developed in the study of the long-range elasticity and anomalous thermoelastic properties of rubber. The following discussion is taken essentially from the excellent presentation of Guth, contained in a monograph by the American Association for the Advancement of Science [37].

Most materials we deal with have a low range of elastic behavior—say 0.2 percent. This is consistent with the short distances over which molecular forces operate. As is well known, rubberlike compounds show elasticity over a much greater range. This is the first general characteristic of rubberlike behavior.

A rod of metal, or almost any other solid, elongates when heated, and has less tensile strength. The forces between the atoms or molecules are weakened by the increase of thermal motion. Rubberlike compounds, on the other hand, contract under a given load when heated, or, analogously, require a greater force to hold them at a given extension. Materials which show the latter type of behavior are selenium and sulfur, natural and synthetic rubbers and, under special conditions, cellulose esters and proteins. These materials are dissimilar in many respects; the only common characteristic is that they are all composed of flexible long-chain molecules.

The flexible chain molecules of unstretched rubber are coiled up in disorder, the most probable state. Anyone who has tossed a piece of string into a catchall drawer can realize the truth of this statement. When rubber is stretched, the chains are necessarily stretched and thus brought into an artificial state of order which is opposed by the internal mobility,

<sup>\*</sup> The mention of commercial products does not imply that they are endorsed or recommended by the Department of Agriculture over others of a similar nature not mentioned.

of internal Brownian motion. Guth and coworkers showed that these considerations can be extended to the three-dimensional network structure of vulcanized rubber. Flexibility of chains persists in the network of cross-bonded chains. This last point is of importance when we extend these considerations to textile fibers.

It is only reasonable that the tensile properties of any actual high-polymer system should exhibit both types of forces—that is, molecular interaction and internal mobility or entropy effects. Thermodynamical analysis of the forces involved shows there is a close analogy with van der Waals' equation, with its two terms due to interaction and kinetic energy. Where F is the total force,  $F_u$  is the internal energy term, and  $F_s$  the kinetic mobility or entropy term:

$$F = F_n + F_s;$$

$$F_n = \left(\frac{\delta U}{\delta L}\right)_T;$$

$$F_s = -T\left(\frac{\delta S}{\delta L}\right)_T = T\left(\frac{\delta F}{\delta T}\right)_L;$$

A plot of stress vs. temperature for various elongations gives the second term  $(F_*)$  directly,\* and this type of analysis can be applied to our system. The value of results which indicate the mechanism by which fibers resist extension is immediately apparent.

But first a word of caution. These ideas neglect the flow or relaxation phenomena so characteristic of fibrous materials. As a practical solution, the fiber is allowed to relax in water under the load until an apparent equilibrium load at each value of L and T is reached. Then a second T is imposed, and the load to give L is determined. Secondly, the effects of swelling on such a system are complex, for swelling is a process dependent on temperature. In fact, Lundgren, Stein, Koorn, and O'Connell [49] have recently used a similar type of analysis to study the interactions between regenerated keratin fibers and water-alcohol mixtures. Nevertheless, the data indicate that these effects, over the temperature range studied, are probably small.

Bull [17] applied thermoelastic analysis to keratin, using human hair as a convenient test material. In general agreement with Astbury's [2, 3] concept of

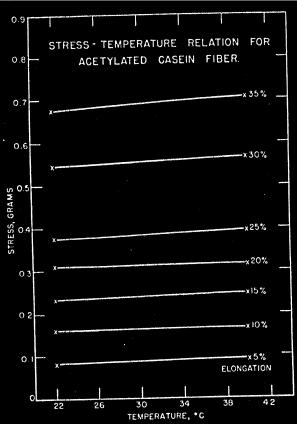


Fig. 1. Stress-temperature relation for acetylated casein fiber.

the  $\alpha$ -keratin- $\beta$ -keratin transformation, he found that hair exists in a rather highly ordered state. Extension is resisted primarily by internal energy forces. At low elongations extension is actually favored by entropy forces; only at high elongations do they resist the stress.

Woods [89] confirmed Bull's results in essence, and showed that as the order in keratin (wool) is destroyed by supercontraction the thermal force becomes of greater importance, and over the initial stages of extension it is large enough to account completely for the load on the fiber. He studied easein fiber, and found that the entropy effect appeared to be much greater than the total load, thus necessitating a negative internal energy. Woods has postulated an internal compression developed by the strong contractile forces as a possible explanation of this anomaly.

We have studied the thermoelastic behavior of our acetylated casein fiber (Figure 1) in the Sookne-Rutherford apparatus [77]. The sample was first relaxed in water at 40°C until the load at definite elongation was reached (about 2 hours). The water

<sup>\*</sup> U= internal energy, L= length, T= temperature in  ${}^{\circ}K$ , S= entropy.  ${\delta F \choose \delta T}_L$  is therefore the slope of the line when F is plotted against T at constant L. (Figure 1); multiplied by T it equals F.  $F_n$  is obtained by subtracting  $F_n$  from F.

bath was then replaced by one at 22°C, and the load necessary to maintain constant length was determined as soon as the system came to equilibrium (less than 1 minute). The cycle can be repeated several times; it is a fully reversible system under these conditions.

The two components of F were then calculated from the slopes of Figure 1, and these were plotted (Figure 2). The internal mobility term,  $F_{s_t}$  is seen to be the dominant force resisting deformation in this fiber up to extensions of 35 percent. This is a demonstration of clearly rubberlike behavior in the wet casein fiber. We did not observe the anomalous hyperentropic effect found by Woods [89].

A third set of considerations, which we cannot directly relate to the internal organization of the fiber but which necessarily is related, was recently developed by Ray [65]. He applied an engineering analysis to the deformations a fiber undergoes in a fabric, and concluded that an individual fiber usually undergoes deformations of not more than 1 or 2 percent. He therefore measured the tensile and torsional properties of fibers within this range, and calculated the work recovery, Young's modulus of elasticity, and the rigidity or torsional modulus. The data were obtained at 70°F and 60 percent relative humidity, and therefore are characteristic of the properties exhibited under normal service conditions.

The results obtained on a series of fibers are of special interest in relation to azlons, for casein fiber closely parallels wool in all these characteristics. We have confirmed the high work recovery, and have demonstrated excellent recovery of length in our continuous-filament casein fiber [64]. Each of the other fibers (viscose, cellulose acetate, fiber A, nylon, silk) differed markedly in one or more factors. All were inferior in work recovery; silk has a high elastic modulus; nylon has a very low rigidity, etc. Further work along this line will be of great interest, for it is probable that fabric qualities such as "hand" and "drape" are the result in large part of the physical properties exhibited at low deformation.

We can therefore conclude that azlon fibers, as typified by casein, are highly elastic. They are composed of flexible chains of undetermined length. The evidence indicates that they possess, at the most, a low degree of internal organization.

#### Amino Acid Composition

Proteins contain some 15 to 20 amino acids in different proportions. Only in especial cases does

one anino acid residue account for more than onefourth the weight. This heterogeneity of composition distinguishes proteins from other high polymers, which often are derived from a single monomer and seldom contain more than two condensing units. This marked difference between proteins and other high polymers has overshadowed the fact that each protein has a distinctive and characteristic chemical composition.

In recent years the differences in composition of wool and silk have received the attention of several authors [39, 70], but consideration has been given primarily to the ease of packing in a crystalline array. Senti has given an especially lucid presentation of this aspect of the problem [70].

We submit the thesis that each protein does have its chemical individuality. This individuality arises from two factors—the number of side groups of the various types and their arrangement. We have little information about the arrangement of the amino acids beyond the vague conviction that there must be a definite pattern. The fact that they are synthesized in a reproducible manner by plants and animals practically precludes the possibility of a random arrangement. But we cannot go further into this matter here, for it is the basic problem of protein chemistry, and despite our meager knowledge would require close analysis of many bits of collateral data.

The most marked progress in the field in recent years has been in the other aspect of the matter—

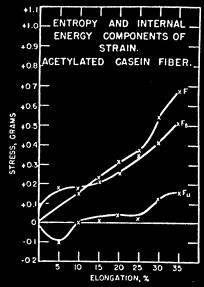


Fig. 2. Entropy and internal energy components of strain—acetylated casein fiber.

TABLE 1. Amino Acid Content of Industrial Proteins

THE STATE OF THE S

Amino acid	Side chain	Casein	Grams ( Keratin (wool)	of amino acid p Fibroin (silk)	per 100 grams Zein (corn)	(soybean)	Arachin (peanut)
Glycine Alanine Valine Leucine Isoleucine Proline Phenylalanine	Hydrogen Methyl Isopropyl Isobutyl Sec-butyl Trimethylene Benzyl	1,90 [73] 3,50 [83] 6,02 [83] 10,55 [83] 5,27 [83] 10,54 [83] 6,46 [83]	6.50 [40] 4.40 [40] 4.72 [40] 11.30 [40] 6.75 [40] 3.75 [40]	42.30 [24] 24.50 [24] 3.20 [24] 0.80 [24] 1.50 [5] 1.50 [24]	9.80 [62] 3.98 [42] 15.40 [12] 4.30 [9] 9.00 [62] 7.60 [25]	0.23 [68] 4.12 [68] 4.50 [7] 8.40 [12] 4.00 [7] 3.94 [8] 5.30 [8]	4.11 [43] 4.85 [42] 8.00 [12] 1.37 [43] 2.60 [43]
Aspartic acid Glutamic acid Ammonia (N)	Carboxymethyl 2-carboxyethyl	6.70 [6] 22.03 [6] 1.41 [66]	7.27 [40] 15.27 [40] 1.40 [66]	2.00 [5] 2.03 [5] 0.00 [24]	3.40 [45] 35.60 [13] 3.64 [62]	3.86 [7] 19.46 [7] 2.02 [7]	5.25 [43] 16.69 [43] 1.67 [43]
Lysine Arginine Histidine	4-amino butyl 3-guanidopropyl Imidazolylmethyl	8,25 [50] 3,94 [50] 3,24 [50]	3.30 [40] 10.40 [40] 0.70 [40]	0.44 [5] 1.05 [5] 0.47 [5]	0.00 [62] 1.60 [85] 0.80 [62]	5.40 [8] 5.80 [8] 2.30 [8]	4.98 [43] 13.51 [43] 1.88 [43]
Serine Threonine Tyrosine	Hydroxymethyl 1-hydroxyethyl Para-hydroxybenzyl	5.87 [66] 4.53 [66] 6.28 [47]	9.41 [40] 6.76 [40] 5.80 [40]	16.24 [66] 1.50 [24] 10.60 [24]	1.00 [69] 2.20 [10] 5.90 [31]	6.00 [26] 4.00 [7] 4.30 [7]	5.20 [14] 2.56 [14] 5.68 [15]
Cystine * Methionine Tryptophane	Bis-thiomethyl 2-methylthioethyl Indolylmethyl	0.40 [44] 3.50 [44] 1.37 [22]	12.72 [40] 0.71 [40] 0.70 [40]	2.59 [5] 1.00 [5]	0.91 [87] 2.30 [69] 0.20 [69]	1.00 [7] 2.00 [7] 1.50 [8]	1.51 [14 0.67 [14 0.68 [15
Total		110.35	110.46	111.72	103.99	86.11	79.54

<sup>\*</sup> Cystine + cysteine.

analyses for the amino acids. As recently as 1940 Vickery [86] divided the amino acids into three groups—those about which our information was little more than qualitative; those for which methods of a considerable degree of probable accuracy had been proposed but had not been widely applied; and,

lastly, those for which existing analytical methods appeared to be satisfactory and had been widely applied. There were 8, 6, and 9, respectively, in these groups. Since then, whole series of new chemical and microbiological methods [1, 11, 20, 74, 75, 78] have been developed and widely applied, so that

TABLE II. Distribution of Amino Acid Residues in Proteins								
	Casein	Moles amin Keratin (wool)	o acid residue Fibroin (silk)	e per 105 gra Zein (corn)	ms of protein Glycinin (soybean)	Arachin (peanut)		
Nonpolar groups (Gly, Ala, Val, Leu, Heu, Phe, Pro)	364	344	883	419	248	176		
Polar groups	101	100	0	260	144	119		
Amide	99	59	29	8	17	34		
Free COOII (diCOOII-amide	56	23	3	0	37	- 34		
Amino (Lys)	44	65	. 9	15	48	90		
Other basic (Arg, His)	9.1	147	168	28	91	72		
Aliphatic OH (Ser, Thr)	35	32	59	30	24	. 31		
Phenolic OH (Tyr)	27	118	22	26	28	20		
Others (Cys ½, Met, Try) Total	456	544	290	367	389	400		
Total groups	820	888	1,173	786	637	576		
% Nonpolar	45	40	75	55	40	. 30		
% Polar	55	60	25	4.5	60	70		

today essentially complete analyses of many proteins are being carried out [11, 16]. As yet the industrial proteins have been given less attention than the relatively pure and homogeneous proteins such as lactoglobulin of whey, serum albumin, and biologically active enzymes, toxins, antitoxins, and hormones. Casein and its two major components, and  $\beta$ -casein, are being carefully investigated in our laboratory [36]. The major seed proteins will probably be studied in the near future.

We have compiled the analytical data on wool, silk, casein, and zein; in each of these proteins the values are good enough for the purpose at hand. The data on wool are taken primarily from the recent compilation of Harris, Mizell, and Fourt [40]. Amino acid analyses of the proteins of soybean, peanut, and cottonseed are not satisfactory. Only about three-quarters of the compositions of these proteins have been accounted for, and new analyses by current methods are greatly needed.

In order that we may consider the amino acid composition from the standpoint of the frequency of reactive groups, we have summarized the structures of the amino acids and the data on each protein, as shown in Tables I and II. In Table, I the amino acids are grouped according to the nature of the side group R,

The same data are summarized and expressed as moles per 100,000 grams in Table II. The average residue weight depends on the amino acid distribution in the protein, and ranges from 78 for silk [24] to 119 for casein. The total number of residues per 100,000 grams therefore would be about 1,280 for silk and 920 for casein, if the amino acid content were completely accounted for.

The nonpolar residues vary in bulk, but all are paraffinic or aromatic. Silk and zein contain 75 and 55 percent, respectively, of these nonreactive groups on a molar basis, and the other proteins contain less than 50 percent. It is probable, however, that more satisfactory analyses of arachin and glycinin would show that the nonpolar amino acids actually make up about half the total in these proteins.

The amide group, derived from the second carboxyl group of aspartic and glutamic acid; is the most frequent polar substituent in the plant proteins. Every third residue in zein and every eighth one in casein is an amide. Wool contains about the same number of amide groups as casein, but has a definitely lower content of free carboxy groups. It is probable that a portion of the amide groups of zein could be hydrolyzed selectively, as indicated by the results of Steinhardt and Fugitt [79], so that the relative proportion of carboxy and carboxy amide groups would approach that of wool rather closely. Silk is anomalous in that all its side-chain carboxy groups are free.

The free amino group of lysine is possibly the most reactive side group in a protein. Amino groups are also contributed by the free α-amino group at the end of the protein chain. Since the contribution of these on a weight basis is small, it is omitted in this tabulation (cf. page 432). The lysine content ranges from zero in zein to 56 moles/105 grams (approximately every fifteenth residue) in casein. Mellon, Korn, and Hoover [54] have shown that the amino group of casein, which constitutes less than 1 percent of the protein on a weight basis. accounts for 25-30 percent of the hygroscopic water sorption. In a second paper it was found that the hysteresis in sorption was not related to this sorption on the amino group, but occurred on other watersorbing centers [55]. McMeekin has correlated the flexibility of formaldehyde-hardened bristle fibers with the lysine content of the proteins [53]. The strongly basic guanidino group of arginine and the imidazole group of histidine are relatively abundant in the proteins considered except for silk and zein, in which they are only about 1 percent of the total.

The relative frequency of aliphatic hydroxyl groups in all the proteins considered, except zein, is high. These hydroxyl groups arise from serine and threo-nine. The existence of serine phosphoric acid (phosphoric acid ester of the hydroxyl group) in cascin has been demonstrated, and, assuming that all the bound phosphorus of cascin (1.6 percent P<sub>2</sub>O<sub>5</sub> [88]) occurs as an aliphatic ester, 23 of the 94 hydroxyl groups are conjugated in this manner. The presence of similarly bound phosphorus in the seed proteins has not been demonstrated.

The amount of phenolic hydroxyl group of tyrosine is remarkably constant in these proteins, about 30 residues per 10° grams occurring in all except silk, which has about twice as many. Tyrosine is especially interesting in that the photochemical

degradation of silk is related to its tyrosine content [67]. Moreover, the solubility of zein in water at pH 11.4 to 12.5 is thought to be due to the dissociation of the phenolic hydroxyl group [60]. Cystine, methionine, and tryptophane occur in rather small amounts in the seed proteins and casein, totaling no more than 30 residues for the three. In wool, in which the cystine linkages play such an important part, there are 106 such residues out of 888.

No discussion of the polar groups of proteins would be complete if the peptide linkage were not included. It is the most frequently occurring group, one existing for each amino acid in the chain. The polar nature of this group is of importance in solubility phenomena [23], and we have found that a large part of the vapor-phase sorption of water by proteins (45–90 percent) occurs on the peptide linkage [56]. Moreover, the formation of hydrogen bonds between imino hydrogen and carbonyl oxygen of peptide linkages in adjacent chains is thought to be of major importance in stabilizing the native structure of proteins. The only chemical reaction which the peptide linkage can be expected to undergo in fiber production is hydrolysis, and that we make every effort to prevent.

A complete discussion of these data from the standpoint of each individual protein would be redundant, but certain conclusions are brought out by comparisons. When compared on the basis of this tabulation, casein and wool show a rather marked similarity. Casein is a highly polar protein with a large number of free carboxy and amino groups. Wool has a high content of guanidino groups, making the total of basic groups high. The noteworthy difference, to which attention has often been called, is the high cystine content of wool and the extremely low cystine content of casein.

Zein has a distinctive composition. The amide linkage is the only polar group which occurs in large amount. It is almost devoid of acid and basic groups. It has therefore a negligible solubility in water and is soluble in polar organic solvents. Swallen and Danehy [82] have given an excellent discussion of the effect of the various polar groups and of nonpolar solvents on the solubility of zein in mixed solvents. Swallen [81] and Manley and Evans [51] established the solvent properties of an extensive series of binary solvents, all of which contained an alcohol as one component. Later Evans and Manley [29] investigated ternary systems based primarily on water, an alcohol, and an aldehyde, ketone, or nitroparaffin.

In these systems the solvent power is correlated with the polar/nonpolar nature of the compounds and their ability to act as hydrogen donors in forming hydrogen bonds. An extensive study of the structure and enzymatic splitting of zein has been carried out by Laine [46].

83.) (1) (1) (編:論新報告 20. 日本)

Silk fibroin is also definitely nonpolar in composition. The only point brought out by the compilation which has not been emphasized before is the high content of aliphatic hydroxyl groups; about every seventh residue has such a substituent. Coleman and Howitt in an outstanding paper [24] have recently reported new chemical and physical data on the structure of silk.

The incompleteness of data on soybean and peanut proteins invalidates any detailed analyses of their amino acid compositions. The nonpolar amino acids were the most difficult to determine before the advent of microbiological methods, and probably the values for this group will be increased when better analyses are available.

In summary, we wish to state again the thesis that each protein has its chemical individuality which arises in the kind and number of its side groups. This fact must be taken into account when we consider the reactions used in producing fiber from industrial proteins. Amino acid determinations of the seed proteins of soybean, peanut, and cottonseed are needed.

#### Chemical Reactions in Fiber Production

The polar side groups are undoubtedly the primary sites of chemical reaction in the intact protein. There are many possible chemical reactions which can be carried out on these various groups. Two recent reviews of the reactivity of the various groups in native proteins each list over 300 references [41, 61]; hence a thorough review of the subject here is not possible. Therefore, only certain designated reactions which relate to fiber production and properties will be discussed.

Reaction with formaldehyde to produce a tougher, more flexible material is an essential step in the production of protein fibers. The chemical nature of this reaction has recently been studied with considerable success, and the results merit our attention. Almost all the nitrogen-containing groups in proteins had been considered as possible sites of reaction with formaldehyde, but, with the exception of the free amino groups, the evidence for that was not

good. Wormell and Kaye [90] and Fraenkel-Conrat, Cooper, and Olcott [33] simultaneously recognized that the amount of formaldehyde combined under acid conditions pointed at participation of the amide group, and they brought forward good supporting evidence. The condensation of amides and formaldehyde to methylol amides and even methylene diamides in strongly alkaline solution was demonstrated by Einhorn [28] in 1905, but the conditions used were far from those commonly used in fiber manufacture (Reaction Scheme I).

Fraenkel-Conrat and Olcott have extended their study of the reaction of formaldehyde with proteins to include other possible sites of reaction. They have made excellent use of model compounds, and have evidence of a number of interesting reactions. The complete evaluation of these studies is difficult, for pH, temperature, time, and concentration all affect any particular experiment.

The guanidino group of arginine reacts with formaldehyde under conditions no more severe than those used in treating protein fiber (Reaction Scheme

- (1) Highly acid conditions, Wormell and Kaye, Fraenkel-Conrat, Cooper, and Olcott
- (1) and (2) Highly alkaline conditions, Einhorn

Reaction Scheme I

Guanidino Group + Formaldehyde

H
N
H
N

-CH<sub>2</sub>-N-C-NH<sub>2</sub> + CH<sub>2</sub>O 
$$\longrightarrow$$
 -CH<sub>2</sub>-N-C-N-CH<sub>2</sub>OH

70°C, pH 3 to 7.6, 4 days

Reaction Scheme II

Amine, CH2O, Amide

pH 3/8, room temperature, 1 to 4 days

Reaction Scheme 111

II). In 4 days, at 70°C and pH 3.0 to 7.6, salmine, a protamine containing about 89 percent of its nitrogen as arginine, combined with more than 1 mole of formaldehyde per guanidyl residue [34].

In an interesting extension of these studies, the reactions of the three-component system amide, formaldehyde, and amine (Reaction Scheme III) have been studied [35]. The rapid reaction in the pII range of 3-8 is the condensation of formaldehyde with primary or secondary amines to form the methylol compound, which then condenses with an amide group. By adding two methylol groups, animonia or primary amines can condense with two amide groups and thus produce cross-links. The amino methylols also condense with guanidino residues. The peptide bond was not reactive in this system, nor did the amido methylol condense with amino groups under the conditions used.

Direct chemical evidence of the reaction products when wool is treated with formaldehyde to produce shrinkproofness and alkali-resistance has been recently obtained. Middlebrook and Phillips [58] have used chromatographic analysis to demonstrate the production of combined thiazolidine-4-carboxylic acid.

after reaction at pH 1.0, 6.7, and 10.0 at 70–100°C. This reaction is of particular interest, for apparently formaldehyde or its methylol has reacted with the nitrogen of the peptide linkage under neutral and acid conditions. Substitution upon this nitrogen has often been postulated, but except for the biuret reaction, which is carried out in highly alkaline solution, evidence for such reactions has not been obtained.

Wool in which a portion of the cystine had been reduced to cysteine reacted with formaldehyde in two different reactions—one to form combined djenkolic acid, the compound produced when two cysteine residues are combined through a methylene bridge—S—CH<sub>2</sub>—S—; the second, that which gave combined thiazolidine-4-carboxylic acid as described above, to give this same compound when it reacted with formaldehyde after reduction to cysteinyl residues.

These studies are of great interest from the standpoint of the protein-formaldehyde reaction, but there is one serious drawback. The reactions discussed so far are reversible; on heating in dilute acid, formaldehyde is evolved, and the groups are set free. The latter conditions are those used in dyeing protein fiber, and all who work in this field have had the experience of seeing a promising fiber go to pieces in the dye bath.

However, nonreversible condensations with formaldehyde also occur [80]. Tryptophane combines with formaldehyde in a stable linkage [32] and with amino methylols, apparently through the >NH group of the ring [35]. The reactive C-H groups of phenols and imidazoles also condense with aminomethylols to produce linkages resistant to acid hydrolysis. These reactions take place with model compounds in neutral or slightly acid conditions, and it appears that the same reactions occur in the reaction of formaldehyde with proteins. The complete evaluation of these reactions will require further extensive study.

Boil stability produced in peanut protein fiber by formaldehyde reaction under drastic conditions of temperature and pH has been demonstrated [30], and we have produced a similar though less pronounced effect on casein fiber. Thus the two lines of attack on the problem—the direct frontal attack and the investigative one—are giving concordant results. It appears probable that this important reaction in protein-fiber chemistry will soon be pretty well understood.

Over the past 5 years we have investigated the effect of a large number of agents on the properties of casein fiber. After the fibers were boiled for an hour in a blank dye bath the results were evaluated primarily on the wet and dry tensile strength and tensile properties. In test after test, none of the reactions appeared to be equal to acetylation with acetic anhydride. This reaction definitely confers boil stability on formaldehyde-hardened fiber. This is a curious result, for it is a monofunctional reagent which has no possibility of producing cross-linkages.

We recently reported results of a silicone treatment of casein fiber [64]. Both wet and dry tensile properties were improved, yet the vapor-phase water take-up was not appreciably affected. We shall not discuss this reaction now, nor the many other stabilizing reactions which we have investigated, for the chemical reactions involved are not well understood. Many of them deserve serious

TAI	BLE III. End-Grou	P DETERMINAT	ions on Puri	FIED PROTEINS		
eta-lactoglobulin	Molecular weight	Total amino N (%)	Epsilon amino N (%)	Alpha amino N (%)	Chain weight	Chains per molecule
practogioniim	35,050 [18] 35,400 [72]	1.260 1.258	1.093	0.167	8,400	4
Bovine serum albumin	70,000 [11]	1.265	1.189	0.076	18,400	4

study, and doubtless the most promising will receive it.

A similar analysis of the participation of the polar side groups in reactions with various inorganic agents is greatly needed. Thus the mode of combination of aluminum, zinc, chromium, and other ions is not at all well understood. The analogy with the use of these metals in tanning leather comes to mind, but there appears to be little clear-cut evidence as to the mechanism of those reactions.

### Chain Length by End-Group Determination

The molecular weights of purified proteins are of considerable interest from the standpoint of protein structure, and many have been determined. number of cases it has been shown that these molecules in solution can be subdivided by such mild treatment as adding urea, surface-active agents, etc. These agents are believed to break hydrogen bonds which hold native protein structures together. Or these submolecules may be held together by normal covalent bonds. The only bond of this type which has been definitely established is the disulfide bond of cystine. There are other possibilities, such as ester linkage between the carboxy and hydroxy side chains, or peptide-like linkages between carboxy and amino side chains. At present there is no evidence of the existence of these hypothesized linkages in proteins but rather good contributory evidence of their nonexistence. We therefore consider the amino acids to be bound by normal peptide linkages into chains, which may or may not be cross-linked.

How long are these chains? In every field of high-polymer study, chain length has been shown to be of major importance in its effect on physical properties. We have been studying this problem for several years, and do not have a completely satisfactory method as yet. The possibility of cystine cross-linkages makes physical methods such as viscosity, streaming birefringence, and the like of questionable applicability. A chemical method must depend on determination of the free primary carboxyl group

at one end of the chain or the free  $\alpha$ -amino group at the other end. The best results we have are from determination of the free  $\alpha$ -amino nitrogen by difference between the total free amino nitrogen and the epsilon-amino nitrogen calculated from lysine content. Lysine is the only amino acid in these proteins which contributes a second amino group. Rees [66] and Desnuelle and Antonin [26] have shown that hydroxylysine, reported by Van Slyke and associates [84], as a constituent in very low amounts of various proteins, is not present in the proteins which we are studying. Hanke's manometric lysine decarboxylase method [38] has given us excellent results. Our enzyme preparations had a slight arginase content, and corrections for this activity were necessary. Total amino nitrogen was determined by the procedure of Doherty and Ogg [27], which was developed in our laboratory for the analysis of insoluble proteins, but is applicable to soluble proteins as well. The time course of the reaction is determined by successive removal of the gas evolved. After some time-30 minutes to 2 hours-the rate of evolution becomes constant at a low value and remains essentially constant for several hours. Extrapolation of the linear portion back to zero time gives the total amino nitrogen with considerably greater certainty than can be obtained from any arbitrary time of reaction. The free terminal amino nitrogen, however, is the relatively small difference between the two values. The consistency of our results is better than might be expected from such a determination by difference. In Table III are reported determinations on  $\beta$ -lactoglobulin and bovine serum (plasma) albumin. These are purified proteins which have been characterized by a variety of physical and analytical methods. The correlation of our data with the physically determined molecular weights indicates 4 chains per molecule for each protein. The data on  $\beta$ -lactoglobulin agree with those of Brand [11] rather well; he obtained 4.6 units per molecular weight of 42,000. He reported 10 free a-amino groups per mole of 70,000 for bovine serum albumin. Our total amino nitrogen

	Total amino N (%)	Epsilon amino N (%)	Alpha amino N (%)	Chain weight		
$\alpha$ -casein	1,00 0,99	0.853 0.855	$0.147 \\ 0.135$	9,500 10,400	10,000	
$\beta$ -casein	0.725 0.715	0.626 0.619	0.099	14,100 \\ 14,500 \}	14,300	
Casein, purified from raw milk	0.94 0.92	0.778 0.782	0.162 0.138	8,600} 10,100}	9,400	
Casein, purified commercial	0.929 0.928	$0.762 \\ 0.762$	0.167 0.166	8,400) 8,400∫	8,400	

value is appreciably lower than the value (1.38 percent) which he reported, and our lysine determination agrees with his. We, therefore, obtain a chain weight of 18,400 and consequently 4 units per mole of 70,000. It must be emphasized that the chain weights determined by these procedures are determined on a weight basis, and are independent of the molecular weight of the protein.

The method has been applied to casein and aand  $\beta$ -casein, its two major components. Purified casein is composed of about 70 percent a- and 30 percent  $\beta$ -casein. A chain length of 10,000 and 14,-000, respectively, was found (Table IV). The purified casein gave a value equal to a chain length of 9,400, obviously not in complete agreement with the value calculated from its two components. A second sample, purified from commercial casein, gave a value of 8,400. The difference between these values of 8,400 and 9,400 is probably not significant, and a value of about 9,000 is indicated for casein. Probably the most reliable molecular weight for comparison is Burk and Greenberg's [19] value of 33,600, determined from the osmotic pressure in urea solution. Again 4 chains per molecule are indicated, but the lack of homogeneity of casein makes such a calculation of questionable value.

We have also applied these method to various seed proteins and to silk fibroin, but the data are only indicative at best. Zein and soybean protein are reported in Table V. The zein was purified from a commercial sample by the procedure of Mason and Palmer [52]. The soybean protein was a commercial protein preparation which may have been contaminated with low-molecular-weight impurities. However, it is probably characteristic of the grade used

TABLE V. End-Group Determinations on Seed Proteins

	Total amino N (%)	Epsilon amino N (%)	Alpha amino N (%)	Chain weight
Zein	0.060 0.061	0.00	0.060 0.061	23,000
Soybean protein	$0.578 \\ 0.562$	$0.419 \\ 0.417$	0.159 0.145	8,800 9,600

in much of the experimental spinning work on this protein.

Zein has been reported to be free of lysine, and we have confirmed this result. Demonstration of the complete absence of an amino acid is difficult, however, and there may be a trace of lysine present as a contaminant. The possible error from this source probably is not much greater than the error in the lysine determinations of those proteins which have an appreciable content of lysine.

Silk fibroin has been difficult to analyze because it is low in lysine, necessitating a large aliquot for the determination. Our data indicate a rather high chain weight, at least as high as the value of 33,000 obtained by Coleman and Howitt [24] by osmotic-pressure measurements.

The results reported, although not completely satisfactory, indicate that the major industrial proteins have a rather low chain length or degree of polymerization (D.P.). The average amino acid residue weights, calculated from the amino acid compositions, range from about 100 to 120. Thus a chain weight of 10,000 corresponds to a D.P. of 80-100. This is about the minimum for satisfactory physical properties proposed by Carothers and Van Natta [21] for polyesters and by Sookne and Harris [76] for cellulose acetate. The minimum D.P. for acceptable protein fiber is difficult to estimate, for it is possible that the interaction of both the polar backbone and the polar side groups may affect it markedly, and a lower D.P. than that in the analogs discussed above may give good fiber properties. This point has been previously considered by Senti, Copley, and Nutting [71].

It must be emphasized that as yet the chain-weight data reported here can be interpreted only with caution. The length of the chain must be of importance in the physical properties of the proteins as solid bodies, but the probability of cross-linkages and

specific modes of folding make correlations of chain length with properties of proteins in solution dangerous. For example,  $\beta$ -lactoglobulin is relatively symmetric in solution, and casein is asymmetric, yet they have about the same chain weight.

These preliminary results give us hope that we can characterize proteins on the basis of their degree of polymerization and thus make much more satisfactory predictions as to the suitability of a protein for fiber production. It is evident that a complete study of the subject will require intensive work on methods of determination and on preparation of pure proteins for measurement. In seed proteins especially, the latter problem is important. We are working on alternate methods of analyses of end groups and hope to get correlative evidence soon.

#### Summary

The major points to which we have drawn attention in this paper are:

Proteins are composed of flexible long chains. The manner in which these chains are arranged controls, to a large extent, the properties of fibers produced from them. Each protein has its own chemical individuality, arising from the number and arrangement of the various amino acid residues, and the length of the polypeptide chain. These differences between proteins must be considered in producing fibers from them.

We believe we have demonstrated that research in proteins is extremely active, and that we can expect to describe their behavior much more accurately in the near future.

#### Literature Cited

- 1. Archibald, R. M., Ann. New York Acad. Sci. 47,
- 2. Astbury, W. T., "Fundamentals of Fibre Structure," London, Eng., Oxford University Press, 1933.
- Astbury, W. T., J. Roy. Soc. Arts 93, 613 (1945).
   Astbury, W. T., Dickinson, S., and Bailey, K., Biochem. J. 29, 2351 (1935)
- 5. Baernstein, Harry D., J. Biol. Chem. 122, 781 (1937)
- Bailey, K., Chibnall, A. C., Rees, M. W., and Williams, E. F., Biochem. J. 37, 360 (1943).
- Block, Richard J., and Bolling, Diana, Arch. Biochem. 3, 217 (1943).
   Block, Richard J., and Bolling, Diana, J. Am. Dietet. Assoc. 20, 69 (1944).
- 9. Block, Richard J., and Bolling, Diana, "The Amino Acid Composition of Foods," Springfield, III., Charles C. Thomas, 1945.

- 10. Borchers, Raymond, Totter, John R., and Berg, Clarence P., J. Biol. Chem. 142, 697 (1942).
- 11. Brand, E., Ann. New York Acad. Sci. 47, 187 (1946).
- Brand, E., Ryan, F. J., and Diskant, E. M., J. Am. Chem. Soc. 67, 1532 (1945).
   Brazier, M. A. B., Biochem. J. 24, 1188 (1930).
   Brown, W. L., J. Biol. Chem. 142, 299 (1942).
   Brown, W. L., J. Biol. Chem. 154, 57 (1944).
   Buehler, H. J., Schantz, E. J., and Lamanna, C., J. Riol. Chem. 169, 295 (1947).

- Biol. Chem. 169, 295 (1947).
- 17. Bull, H. B., J. Am. Chem. Soc. 67, 533 (1945).
- 18. Bull, H. B., and Currie, B. T., J. Am. Chem. Soc.
- 68, 742 (1946).
  19. Burk, N. F., and Greenberg, D. M., J. Biol. Chem. 87, 197 (1930).
- 20. Cannan, R. K., Ann. New York Acad. Sci. 47, 135 (1946).
- 21. Carothers, W. H., and Van Natta, F. J., J. Am. Chem. Soc. 55, 4714 (1933).
- 22. Chibnall, A. C., Proc. Roy. Soc. B, 131, 136 (1942). 23. Cohn, E. J., and Edsall, J. T., "Proteins, Amino Acids and Peptides" (A.C.S. Monograph No. 90), New York, Reinhold Publishing Corp., 1943.
- 24. Coleman, D., and Howitt, F. O., Proc. Roy. Soc. A, 190, 145 (1946).
- 25. Dakin, H. D., Z. Physiol. Chem. 130, 159 (1923).
- 26. Desnuelle, P., and Antonin, S., Biochim. et Biophys. Acta 1, 50 (1947).
- 27. Doherty, David G., and Ogg, Clyde L., Ind. Eng. Chem., Anal. Ed. 15, 751 (1943).
- 28. Einhorn, A., Ann. 343, 207 (1905).
- 29. Evans, C. D., and Manley, R. H., Ind. Eng. Chem. 36, 408 (1944).
- 30. Fieldsend, E. C., and Boyes, W. H. D., British Patent 513,910 (Oct. 25, 1939).
- 31. Folin, O., and Ciocalten, V., J. Biol. Chem. 73, 627
- 32. Fraenkel-Conrat, H., Brandon, B. A., and Olcott, H. S., J. Biol. Chem. 168, 99 (1947).
- 33. Fraenkel-Conrat, H., Cooper, M., and Olcott, H. S., J. Am. Chem. Soc. 67, 950 (1945).
- Fraenkel-Conrat, H., and Olcott, H. S., J. Am. Chem. Soc. 68, 34 (1946).
- 35. Fraenkel-Conrat, H., and Olcott, H. S. In press. 36. Gordon, W. G., Semmett, W. F., Cable, R. S., and Doherty, D. G., Federation Proc. 6, 255 (1947).
- 37. Guth, E., in "Surface Chemistry," Washington, D. C., Am. Assoc. Advance. Sci., Pub. No. 21, 1943, p. 103.
- 38. Hanke, M., Federation Proc. 5, 137 (1946).
- 39. Harris, M., and Brown, A. E., Textile Research
- JOURNAL 17, 323 (1947).
  40. Harris, M., Mizell, L. R., and Fourt, L., Ind. Eng. Chem. 34, 833 (1942).
- Herriott, R. M., in "Advances in Protein Chemistry," Vol. III, New York, Academic Press, Inc., 1947, p. 169.
   Horn, M. F., Jones, D. B., and Blum, A. E., J. Biol. Chem. 170, 719 (1947).

43. Johns, C. O., and Jones, D. B., J. Biol. Chem. 36, 491 (1918).

44. Kassell, B., and Brand, E., J. Biol. Chem. 125, 435 (1938).

45. Laine, T., Suomen Kemistilehti 12B, 23 (1939).

46. Laine, T., Ann. Acad. Sci. Fennicae, Ser. All, Chimica No. 11, 7 (1944).

47. Lugg, J. W. H., Biochem. J. 32, 2123 (1938).

- 48. Lundgren, H. P., and O'Connell, R. A., Ind. Eng. Chem. 36, 370 (1944).
- 49. Lundgren, H. P., Stein, A. M., Koorn, V. M., and O'Connell, R. A., J. Phys. & Colloid Chem. 52, 180 (1948).
- MacPherson, H. T., Biochem. J. 40, 470 (1946).
   Manley, R. H., and Evans, C. D., Ind. Eng. Chem. 35, 661 (1943).
- 52. Mason, Inez D., and Palmer, Leroy S., J. Biol. Chem. 107, 131 (1934). 53. McMeekin, T. L., A. S. T. M. Bull., No. 125, 19
- (Dec. 1943).
- 54. Mellon, E. F., Korn, A. H., and Hoover, S. R., J. Am. Chem. Soc. 69, 827 (1947).
- Mellon, E. F., Korn, A. H., and Hoover, S. R., J. Am. Chem. Soc. 70, 1144 (1948).
- 56. Mellon, E. F., Korn, A. H., and Hoover, S. R.,
- J. Am. Chem. Soc., July, 1948 (in press).57. Meyer, K. H., and Mark, H., Ber. B 61, 1932 (1928).
- 58. Middlebrook, W. R., and Phillips, H., Biochem. J. 41, 218 (1947)
- 59. Nutting, G. C., Halwer, M., Copley, M. J., and Senti, F. R., Textile Research Journal 16, 599 (1946).
- 60. Ofelt, C. W., and Evans, C. D. In press.
- 61. Olcott, H. S., and Fraenkel-Conrat, H., Chem. Rev. 41, 151 (1947).
- Osborne, T. B., and Liddle, L. M., Am. J. Physiol. 26, 295 (1910).
- 63. Peterson, R. F., Caldwell, T. P., Hipp, N. J., Hellbach, R., and Jackson, R. W., Ind. Eng. Chem. 37, 492 (1945).
- 64. Peterson, R. F., McDowell, R. L., and Hoover, S. R., Paper presented at 112th Meeting, Am. Chem. Soc., New York, Sept., 1947.
- 65. Ray, L. G., Jr., TEXTILE RESEARCH JOURNAL 17, 1 (1947).

- 66. Rees, M. W., Biochem. J. 40, 632 (1946).
- 67. Rutherford, H. A., and Harris, Milton, J. Research Natl. Bur. Standards 27, 81 (1941).
- 68. Sasaki, Shuiku, J. Agr. Chem. Soc. Japan 11, 321
- (1935). 69. Schmidt. Carl L. A., "Chemistry of the Amino Acids and Proteins," Springfield, Ill., Charles C. Thomas, 1938.
- 70. Senti, F. R., Am. Dyestuff Reptr. 36, 230 (1947).
- Senti, F. R., Copley, M. J., and Nutting, G. C., J. Phys. Chem. 49, 192 (1945).
- 72. Senti, F. R., and Warner, R. C., J. Am. Chem. Soc., Aug., 1948 (in press).
- Shankman, S., Camien, M. N., and Dunn, M. S., J. Biol. Chem. 168, 51 (1947).
- 74. Shemin, D., and Foster, G. L., Ann. New York Acad. Sci. 47, 119 (1946).
- 75. Snell, E. E., Ann. New York Acad. Sci. 47, 161 (1946).
- 76. Sookne, A. M., and Harris, M., J. Research Natl.
- Bur. Standards 30, 1 (1943). 77. Sookne, A. M., and Rutherford, H. A., J. Research
- Natl. Bur. Standards 31, 25 (1943). 78. Stein, W. H., and Moore, S., Ann. New York Acad.
- Sci. 47, 95 (1946).
- 79. Steinhardt, J., and Fugitt, C. H., J. Research Natl. Bur. Standards 29, 315 (1942).
- 80. Swain, A. P., Kokes, E. L., Hipp, N. J., Wood, J. L., and Jackson, R. W., Ind. Eng. Chem. 40, 465 (1948).
- 81. Swallen, L. C., Ind. Eng. Chem. 33, 394 (1941).
- 82. Swallen, L. C., and Danehy, J. P., "Zein," in Alexander, J., "Colloid Chemistry," Vol. VI, New York, Reinhold Publishing Corp., 1946, p. 1140.
- 83. Tristram, G. R., Biochem, J. 40, 721 (1946).
- 84. Van Slyke, D. D., Hiller, A., and MacFadyen, D. A., J. Biol. Chem. 141, 681 (1941).
- 85. Vickery, H. B., J. Biol. Chem. 132, 325 (1940).
- 86. Vickery, H. B., Ann. New York Acad. Sci. 41, 87 (1941).
- 87. Vickery, H. B., and White, A., J. Biol. Chem. 99, 701 (1933).
- 88. Warner, R. C., J. Am. Chem. Soc. 66, 1725 (1944).
- 89. Woods, H. J., J. Colloid Sci. 1, 407 (1946). 90. Wormell, R. L., and Kaye, M. A. G., J. Soc. Chem. Ind. 64, 75 (1945).

(Manuscript received April 27, 1948.)

# "Constitutional Factors in the Production of Artificial Protein Fiber,"\* by Sam R. Hoover, Elsie L. Kokes, and Robert F. Peterson

## The Entropy Contribution to the Load in Stretched Casein

In the above paper the authors discuss some results of mine relating to the entropy and internal energy contributions to the load on stretched casein fibers (J. Colloid Sci. 1, 407 (1946)). With their notation the total load, F, may be written as the sum of  $F_n$ , the entropy term, and  $F_n$ , the internal energy term; I found that at low extensions (up to 25% in some fibers)  $F_u$  was negative, becoming positive at higher extensions, and I coined the phrase "hyperentropie" to describe the state of affairs where  $F_* >$ F. Describing experiments of their own on acetylated casein, Hoover, Kokes, and Peterson state (p. 426): "We did not observe the anomalous hyperentropic effect found by Woods." This statement was surprising as I had already noted that in their Figure 2 the authors had plotted a curve for  $F_n$  vs. elongation in which negative values are shown for extensions up to about 10%, and I was congratulating myself on an independent confirmation that such an effect occurs, admittedly over a lower extension range, but this might easily happen in a fiber so notoriously variable as casein.

I have already pointed out (loc. cit.) that in at least one other fiber, nylon, a negative initial value of

TEXTILE RESEARCH JOURNAL 18, 423 (July 1948).

 $F_u$  is observed, but only over an extension range of some 4%. I should like to add that I have results for "elastic" nylon which show the effect for extensions up to 40%.

H. J. Woods Textile Physics Laboratory, The University, Leeds, England

(October 29, 1948)

#### **Authors' Comments**

We appreciate Dr. Woods' comments on our paper. The magnitude of the effect we observed was small. It was relatively consistent in various experiments and may well be a real effect. On the other hand, the small magnitude leaves open the possibility of a systematic error of manipulation or of the instrument. Our main point in this connection was that the calculated entropy load was not several times the total load at extensions up to 20% (three times the observed tension at 10% elongation) as Figure 6 of Dr. Woods' paper indicates.

SAM R. HOOVER
ELSIÉ L. KOKES
ROBERT F. PETERSON
Eastern Regional Research Laboratory,
Philadelphia 18, Pa.

(November 23, 1948)